Structural Biology

Deutsche Ausgabe: DOI: 10.1002/ange.201509150
Internationale Ausgabe: DOI: 10.1002/anie.201509150

N-Linked Glycans of Chloroviruses Sharing a Core Architecture without Precedent

Cristina De Castro, Immacolata Speciale, Garry Duncan, David D. Dunigan, Irina Agarkova, Rosa Lanzetta, Luisa Sturiale, Angelo Palmigiano, Domenico Garozzo, Antonio Molinaro, Michela Tonetti, and James L. Van Etten*

Abstract: *N-glycosylation is a fundamental modification of proteins and exists in the three domains of life and in some viruses, including the chloroviruses, for which a new type of core N-glycan is herein described. This N-glycan core structure, common to all chloroviruses, is a pentasaccharide with a β-glucose linked to an asparagine residue which is not located in the typical sequon N-X-T/S. The glucose is linked to a terminal xylose unit and a hyperbranched fucose, which is in turn substituted with a terminal galactose and a second xylose residue. The third position of the fucose unit is always linked to a rhamnose, which is a semiconserved element because its absolute configuration is virus-dependent. Additional decorations occur on this core N-glycan and represent a molecular signature for each chlorovirus.*

Protein N-glycosylation is a fundamental post-translational modification which occurs in all domains of life, eukarya, archaea, and bacteria,[1] and also in many viruses, such as rhabdoviruses, herpesviruses, poxviruses, and paramyxoviruses.[2] These viruses infect eukaryotic organisms and use the host-encoded machinery to add and remove sugars from virus glycoproteins. Consequently, *N*-acetylglucosamine is the sugar directly linked to the Asn residue in these viral glycoproteins and the glycan portion resembles host glycans. Therefore, the only way to alter glycosylation of virus proteins is to either grow the virus in a different host[3] or have a mutation which changes the viral protein such that it alters glycosylation.

Chloroviruses (family *Phycodnaviridae*) infect eukaryotic algae but differ from this paradigm because they encode most, if not all, of the machinery to glycosylate their major capsid proteins (MCP).[4] This genus includes large (190 nm in diameter) icosahedral, plaque-forming viruses with an internal lipid membrane and dsDNA genomes encoding as many as 400 proteins.[5] We have recently shown that Vp54, the MCP of *Paramecium bursaria* chlorella virus (PBCV-1) contains four N-linked glycans,[6] with structures that do not resemble any other reported for either bacteria, archaea, or eukarya,[7] including unicellular algae.[8]

The PBCV-1 Vp54 glycans are attached to the protein by an Asn β-glucose linkage. The Asn is not located in the typical N-X-T/S consensus sequon. The glycans consist of eight to ten neutral monosaccharide residues, for a total of four glycoforms (Figure 1).[7] These N-glycans have a highly branched architecture with two residues, arabinose and mannose, as nonstoichiometric substituents, along with a hyperbranched fucose, and a dimethylated rhamnose as a capping residue. Prompted by the peculiarity of these motifs, we examined the N-glycans of additional chloroviruses to determine if those described for PBCV-1 were unique, or if their structural features were conserved in the other chloroviruses and represent an unprecedented N-glycosylation pattern.

Five chloroviruses with different host specificities were chosen: NY-2A, like PBCV-1, infects *Chlorella variabilis* NC64A (NC64A viruses), ATCV-1 and TN603 infect *Chlorella heliozoae* (SAG viruses), and MT325 and CVM-1 infect *Micractinium conductrix* Pbi (Pbi viruses). The MCP amino-acid sequence from PBCV-1 was used to identify the MCPs from the five additional chloroviruses (see Figure S1 in the Supporting Information). All of these viruses encode one MCP with a calculated molecular weight lower than that determined experimentally by either ESI-MS (Table 1) or SDS-PAGE (Figure 2).

The ESI-MS data indicated a molecular weight increase of 5.9–12.5% for all the MCPs, except for MT325, for which clean MS data were not obtained, and the SDS-PAGE positive staining for glycoproteins (Figure 2A) suggested that these MCPs were glycosylated. The virus NY-2A was an exception: it encodes two closely related (or paralogue) MCPs, B585L and B617L (see Table S1 in the Supporting Information), both glycosylated as deduced by SDS-PAGE (Figure 2A) and ESI-MS data (Table 1). These two MCPs

[*] Prof. C. De Castro
Department of Agricultural Sciences, University of Napoli
Via Università 100, 80055 Portici (NA) (Italy)
E-mail: decastro@unina.it
Dr. I. Speciale, Prof. R. Lanzetta, Prof. A. Molinaro
Department of Chemical Sciences, University of Napoli
Via Cintia 4, 80126 Napoli (Italy)
Prof. G. Duncan
Department of Biology, Nebraska Wesleyan University
Lincoln, NE (USA)
Dr. D. D. Dunigan, Dr. I. Agarkova, Prof. J. L. Van Etten
Department of Plant Pathology and Nebraska Center for Virology,
University of Nebraska, Lincoln, NE 68583-0900 (USA)
Dr. L. Sturiale, Dr. A. Palmigiano, Prof. D. Garozzo
CNR, Institute for Polymers, Composites and Biomaterials,
95126 Catania (Italy)
Prof. M. Tonetti
Department of Experimental Medicine and Center of Excellence for
Biomedical Research, University of Genova
Viale Benedetto XV/1, 16132 Genova (Italy)
Supporting information and ORCID(s) from the author(s) for this article are available on the WWW under http://dx.doi.org/10.1002/anie.201509150.

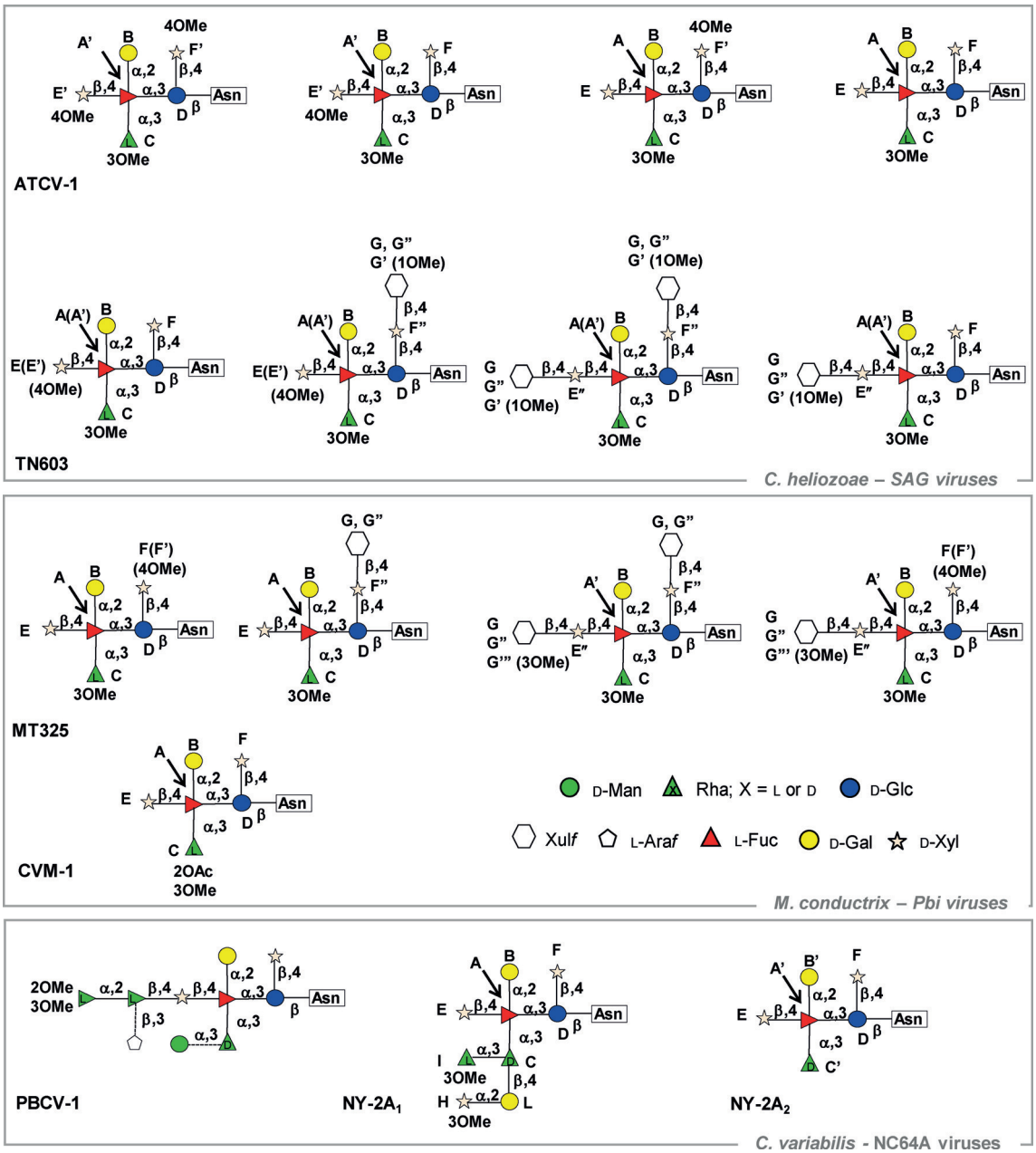

Figure 1. Structures of N-glycans from six chloroviruses. The basic core structure is defined by the residues **A**, **B**, **D**, and **F**, and was conserved between all the chloroviruses. The xylose **F** is named the "proximal xylose" because it is closer than **E** to the peptide backbone; **E** is the "distal xylose". Inclusion of **C** gives the extended conserved core structure. The N-glycan of the prototype chlorovirus PBCV-1 is given for comparison and monosaccharides connected by broken lines are not stoichiometric substituents. Structures from the same virus are arranged in a row, and different viruses are grouped according to host specificity, that is indicated at the right-bottom margin of the grouping box. Letter labels are those used during NMR assignment, and all sugars are in the pyranose form except where specified.

could not be separated and further analysis was performed on the combined pair.

Glycopeptide isolation from each MCP was accomplished as reported (see the Supporting Information for details).[7] Each virus gave one glycopeptide (^{1}H NMR spectrum in Figure 3), except NY-2A for which two different fractions were isolated, NY-2A$_1$ and NY-2A$_2$.

Monosaccharide chemical analysis (see Table S2 in the Supporting Information)[7] identified all the residues (see Figure S2), except for xylulose, a labile residue which degrades during the methanolysis reaction. The structure of each glycopeptide was obtained by combined analysis of the NMR and MALDI data (see the Supporting Information). MALDI MS studies supported NMR results and determined the glycosylation sites for CVM-1 and NY-2A.

Virus CVM-1 had the simplest glycan among all the viruses in this study (Figure 1), and consisted of only six monosaccharides as noted in the HSQC spectrum (see Figure S3). The unit **D** ($\delta = 5.01$ ppm) had the anomeric carbon atom N-linked to the protein as disclosed by its carbon

Table 1: Major capsid protein statistics for six chloroviruses: PBCV-1 (A430L), NY2A (B585L and B617L, two nearly identical paralogs), MT325 (M463L), CVM-1 (548L), TN603.4.2 (325L), and ATCV-1 (Z280L).

Virus (gene code)	% AA identity	MW°[a]	MW	ΔMW (%)	N-glycan MW[b]	Avg. no. glycans	Asn sequon found
PBCV-1 (A430L)	100	48168	53790[c]	5622 (11.7)	1384	4	^{280}NIPG; ^{302}NTGT; ^{399}NTET; ^{406}NTAT
NY-2A (B585L) [d]	96	48359	52641	4282 (8.9)	1348	4 or 3[e]	^{54}NKVS;[f] ^{280}NIPG;[g] ^{302}NTGT,[g] ^{399}NTET;[g] ^{406}NTAT[g]
			51250	2891 (5.9)	880		
NY-2A (B617L) [d]	94	48354	As above	As above	As above	As above	As above and ^{291}NVAT[h]
MT325 (M463L)	78	47652	n.d.	n.d.	1186	n.d.	n.d.
CVM-1 (548L)	79	47622	51247	3625 (7.5)	936	4	^{47}NGSV;[i] ^{279}NLTA;[g] ^{285}NVGY;[g] ^{293}NTAV[g]
TN603.4.2 (325L)	79	48437	54508[j]	6071 (12.5)	1186	5	n.d.
ATCV-1 (Z280L)	78	48421	52827	4406 (9.1)	922	5	n.d.

The ΔMW% represents the net increase in percentage of molecular the weight for each protein, calculated as: (MW−MW°)×100/MW. [a] MCP molecular weight without glycans. [b] For NY-2A the molecular weight of the two oligosaccharides are given. For ATCV-1, TN603, and MT325 the MW of the largest oligosaccharide is listed. [c] Taken from Ref. [12]. [d] Experimental work performed on the mixture of the two NY-2A MCPs, for which two different MWs were detected. Nonasaccharide was at Asn-302, Asn-399, Asn-406; hexasaccharide at Asn-54, Asn-280, and Asn-291. [e] Average number calculated considering two nonasaccharides and two hexasaccharides (major ΔMW) or one nonasaccharide and two hexasaccharides (minor ΔMW). [f] Asn glycosylated but previously predicted or conserved in PBCV-1. [g] Conserved with respect to PBCV-1 (see alignment sequence in Figure S1). [h] Site not predicted and present only in NY2A MCP: B617L. [i] Glycan present in this site in CVM-1 but not in PBCV-1 and NY2A. [j] MCP highly glycosylated, higher MW is given in the table, but these other masses were obtained after spectrum deconvolution: 54363, 54227, 54079, and 53950 Da. n.d. = not determined.

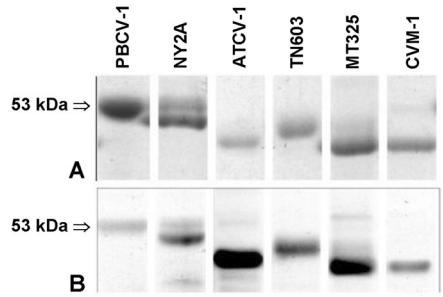

Figure 2. SDS-PAGE of total proteins of the MCPs from chloroviruses PBCV-1, NY-2A, ATCV-1, TN603, CVM-1, and MT325. A) Glycoprotein staining. B) Silver staining. Note that NY-2A has two bands.

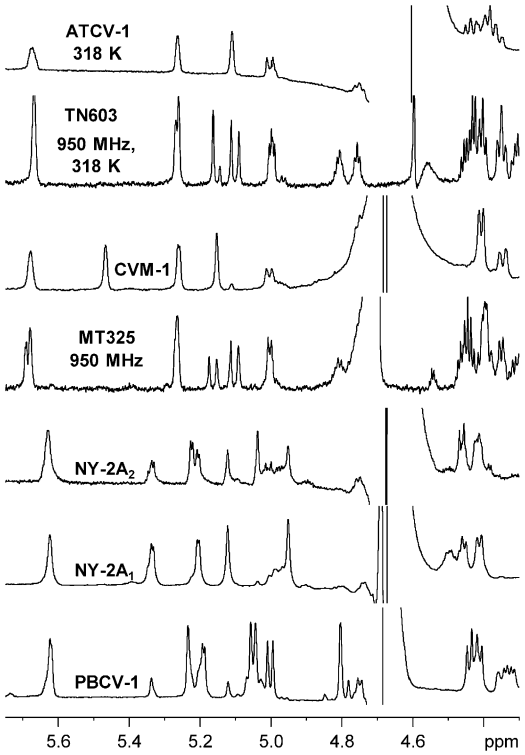

Figure 3. ^{1}H NMR spectra reporting the anomeric region of the glycopeptides obtained from the capsid protein of the different chloroviruses. Where not specified, spectra were recorded at 600 MHz and at 310 K. The spectrum of PBCV-1 is provided as a reference.

chemical shift ($\delta = 80.7$ ppm), and it was a glucose (see Figures S4 and S5 and Table S3), substituted as in PBCV-1 with an α-fucose (**A**) at O3, and a β-xylose **F** at O4. This last xylose is named the proximal xylose because it is closer than the other xylose to the peptide backbone. **A** was fully substituted and it had the α-galactose **B** at O2, α-rhamnose **C** at O3, and finally the distal β-xylose **E** at O4.

Compared to PBCV-1, CVM-1 glycan had some similarities with a few differences (Figure 1): it was not elongated at the distal xylose, and **C** was L-configured instead of D-configured, in addition to being acetylated at O2 and methylated at O3. Consistent with the different absolute configuration of **C**, carbon chemical shifts (see Table S4), of **A** in CVM-1, diverged from those of the equivalent residue of PBCV-1,[7] and such variation fitted with the rules reported by Shashkov et al.[9] MALDI MS confirmed the CVM-1 N-glycan structure and determined the sites of attachment (see Figure S6 and Table 1): Asn-47, Asn-279, Asn-285, and Asn-293. With the exception of Asn-47, all positions were equivalent to those found glycosylated in PBCV-1 (see Figure S1). Asn-47 is the only one found in a typical sequon (NGS; Table 1) and although it is conserved in all MCPs, its glycosylation only occurs in CMV-1.

The ATCV-1 N-glycan (Figures 1 and Figures 3; see Figures S7 and S8 and Table S4) had several similarities to CVM-1. The main differences occurred in **C** and the two xylose units. The anomeric signal of **C** was slightly different

because this unit was methylated at O3, but not acetylated at O2, while inspection of the xylose anomeric region (see Figure S8C) indicated a high degree of complexity. Such complexity originated from a methyl group which was not stoichiometric at O4 of each xylose unit, thus generating four different glycopeptides (Figure 1) as confirmed by MALDI spectrometry (see Figure S9).

TN603 glycopeptides (Figure 3; see Figures S10 and S11 and Table S5) shared with ATCV-1 the complex pattern of xylose residues, which depended on two nonstoichiometric substituents: a methyl group and a xylulose residue. These two appendages were always at O4 of the xylose residues, with the methyl group occurring only at the distal unit while xylulose was at both xylose units. In addition, xylulose was also methylated in a nonstoichiometric fashion at O1, thus increasing the overall variability of this N-glycan as depicted in Figure 1. The great variation in structures observed by NMR spectroscopy was confirmed by MALDI MS analysis (see Figure S12 and detailed descriptions given in the Supporting Information).

The NMR spectra of MT325 (Figure 3; see Figures S13 and S14 and Table S6) resembled those of TN603: divergences related with the nonstoichiometric substitution pattern at O4 of the two xylose residues. The methyl group was only at the proximal xylose, while xylulose was at both xylose residues, with that at the distal xylose further decorated with a methyl group at O3.

As a result, several complex glycoforms were identified for MT325 glycopeptides and supported by MALDI spectrometry (see Figure S15). The most relevant structures are reported in Figure 1.

NY-2A MCPs produced two glycopeptide fractions: interestingly, in NY-2A$_1$ (Figure 3; see Figures S16 and S17 in the Supporting Information) carbon chemical-shift values for **A** (see Table S7) were similar to those found for the equivalent residue in PBCV-1. This agreement resulted from the nature of **C**, a α-rhamnose at O3 of **A**, having the D absolute configuration as in PBCV-1, and not L as for the other viruses. **C** was substituted at both O3 and O4 by α-L-3OMeRha (**I**) and α-D-Gal (**L**), which was in turn substituted at O2 with a α-D-3OMeXyl residue (**H**). Taken together, this information defined the NY-2A$_1$ glycopeptide as a highly branched nonasaccharide (Figure 1).

NY-2A$_2$ NMR spectra (see Figure 3 and Figures S18–S20 and Table S8) contained all the signals assigned in NY-2A$_1$ along with a new anomeric signal at $\delta = 5.04$ ppm, labelled **C′**. This new residue was a nonsubstituted α-rhamnose linked at O3 of the hyperbranched fucose. Thus, NY-2A$_2$ was a mixture of two different glycopeptides, a nonasaccharide and its truncated form (Figure 1). MALDI spectrometry confirmed this information (see Figure S21) and also identified the glycosylation sites of the MCPs, which occurred at the same positions (Asn-280, Asn-302, Asn-399, and Asn-406) as in PBCV-1, along with two additional ones, Asn-54 and Asn-291 (Table 1, and Figure S21).

Glycosylation of NY-2A MCPs is interesting for several reasons. First, Asn-54 is in a region conserved in all MCPs but not glycosylated in PBCV-1 and CVM-1, while Asn-291 is not conserved and only exists in the NY-2A B617L paralogue. Second, glycosylation is site-specific: nonasaccharide occurs at Asn-302, Asn-399, and Asn-406, while the hexasaccharide occurs at Asn-54, Asn-280, and Asn-291 (Table 1). Finally, the presence of five (B585L) or six (B617L) glycosylation sites does not fit with the average number of glycans calculated for these MCPs (Table 1), thus suggesting that not all of the possible sites are always occupied in each paralogue, an issue for future studies.

Hence, solving the N-glycan structures of five additional chloroviruses addresses the question which prompted this study: this is an unprecedented N-glycosylation pattern. All the MCPs from these chloroviruses contain N-linked glycans (Figure 1) which share a unique central core structure (Figure 4). This motif starts with a N-linked glucose and contains a hyperbranched fucose, a distal and a proximal xylose, and a galactose.

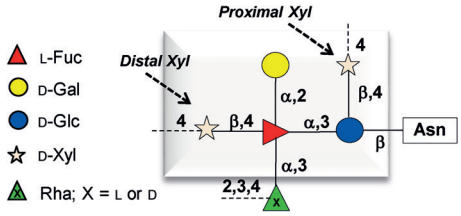

Figure 4. Oligosaccharide core structure common to the N-glycans of the MCPs from six different chloroviruses. The strictly conserved region is made by the five monosaccharides inside the box. The unit outside is always rhamnose and its inclusion in the formula gives the extended form of the conserved core structure. The absolute configuration of this rhamnose is D for NC64A viruses (NY-2A and PBCV-1), otherwise, it is L-configured and methylated at O3. Broken lines indicate positions further decorated: the type and number of appendages is virus specific as detailed in Figure 1.

This basic core motif is extended with a semiconserved element: a rhamnose is always at O3 of the fucose. However, this unit has the D configuration in NC64A viruses, while it is L-configured and methylated at O3 in the SAG and Pbi viruses (Figure 4). Therefore, its inclusion in the basic core structure generates an extended core motif which is group specific. This extended core structure is further decorated with other appendages and the overall oligosaccharide is the signature of each virus (Figure 1). Other relevant features are the N-glycans capped with one or more methyl groups, a modification common to many organisms except mammals,[11] while the glycosylation sites determined in this study for some of the MCPs confirmed that they do not occur in the typical sequon N-X-T/S, with the only exception being Asn-47 in CVM-1.

In conclusion, N-glycans from chloroviruses have a new structural motif which can be considered either as a signature or as a distinctive trait for this class of organisms, thus opening the way to many intriguing questions. For example, how does biosynthesis occur and what are the genes/proteins involved in the process? How is the glucose–asparagine linkage assembled? How does the L,D-rhamnosyl transferase work to equally transfer both L- and D-rhamnose nucleotides to the growing oligosaccharide? But also, how much is this motif

diffused throughout nature and how many non-eukaryotic-like N-glycosylation patterns have been overlooked so far? Are the chloroviruses the only organisms possessing their own N-glycans? Answers to these questions will increase the global knowledge of glycoscience and will open new avenues to understanding such an important matter as the biosynthesis and function of glycoproteins in nature.

Acknowledgments

Work in the Van Etten laboratory was partially supported by National Science Foundation grant EPS-1004094, Stanley Medical Research Institute grant 11R-003, and the National Institute of Health grant P20-RR15635 from the Centers of Biomedical Research Excellence Program of the National Center for Research Resources. G.D. was supported by National Institute for General Medical Science grant P20GM103427. Work in the Garozzo laboratory was partially supported by Stellalucente Trust and the Italian National Council of Research (CNR). C.D.C. and A.M. wish to acknowledge financial support from NMR IR-RMN-THC Fr3050 CNRS for the use of 950 MHz NMR spectrometer.

Keywords: glycoproteins · glycosylations · natural products · structural biology · viruses

How to cite: *Angew. Chem. Int. Ed.* **2016**, *55*, 654–658
Angew. Chem. **2016**, *128*, 664–668

[1] F. Schwarz, M. Aebi, *Curr. Opin. Struct. Biol.* **2011**, *21*, 576–582.
[2] S. J. Flint, L. W. Enquist, V. R. Racaniello, A. M. Skalka, *Principles of Virology, Molecular Biology, Pathogenesis, and Control of Animal Viruses*, ASM Press, Washington, DC, **2004**.
[3] D. J. Vigerust, V. L. Shepherd, *Trends Microbiol.* **2007**, *15*, 211–218.
[4] J. L. Van Etten, J. Gurnon, G. Yanai-Balser, D. Dunigan, M. V. Graves, *Biochim. Biophys. Acta Gen. Subj.* **2010**, *1800*, 152–159.
[5] J. L. Van Etten, D. D. Dunigan, *Trends Plant Sci.* **2012**, *17*, 1–8.
[6] N. Nandhagopal, A. Simpson, J. R. Gurnon, X. Yan, T. S. Baker, M. V. Graves, J. L. Van Etten, M. G. Rossmann, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14758–14763.
[7] C. De Castro, A. Molinaro, F. Piacente, J. R. Gurnon, L. Sturiale, A. Palmigiano, R. Lanzetta, M. Parrilli, D. Garozzo, M. Tonetti, L. J. Van Etten, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 13956–13960.
[8] E. Mathieu-Rivet, M. C. Kiefer-Meyer, G. Vanier, C. Ovide, C. Burel, P. Lerouge, M. Bardor, *Front. Plant Sci.* **2014**, *5*, 1–13.
[9] A. S. Shashkov, G. M. Y. Lipkind, Y. A. Knirel, N. K. Kochetkov, *Magn. Reson. Chem.* **1988**, *26*, 735–747.
[10] P. Stanley, H. Schchter, N. Taniguchi in *Essential of Glycobiology*, 2nd ed. (Eds.: A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanly, C. R. Bertozzi, G. W. Hart, M. E. Etzler), Cold Spring Harbor, NY, **2009**, pp. 101–114.
[11] E. Staudacher, *Biol. Chem.* **2012**, *393*, 675–685.
[12] M. V. Graves, C. T. Bernadt, R. Cerny, J. L. Van Etten, *Virology* **2001**, *285*, 332–345.

Received: September 30, 2015
Published online: November 19, 2015